REDIC: Recommendation of Digital Influencers of Brazilian Artisanal Cheese

REDIC: Recomendação de Influenciadores Digitais do Queijo Artesanal Brasileiro

Nedson D. Soares
Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil
nedson.soares@estudante.ufjf.br

Regina Braga
Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil
regina.braga@ufjf.edu.br

José Maria N. David
Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil
jose.david@ufjf.edu.br

Kennya B. Siqueira
Embrapa Gado de Leite, Juiz de Fora - MG, Brasil
kennya.siqueira@embrapa.br

Thallys S. Nogueira
Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil
thallys.nogueira@estudante.ufjf.br

Emerson W. Campos
Instituto Federal Sudeste de Minas Gerais, Juiz de Fora - MG, Brasil
emerson.campos@colaborador.embrapa.br

Emerson A. P. Moraes
Instituto Federal Sudeste de Minas Gerais, Juiz de Fora - MG, Brasil
emerson.moraes@ifsudestemg.edu.br

Priscila V.Z.C. Goliatt
Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil
capriles@ice.ufjf.br

ABSTRACT

The advancement of information technology makes social media networks increasingly gain popularity and insertion in daily life aspects. Thus, the analysis of people's opinions and habits is essential for many companies' modernization and survival. On social networks, people generally share their views and visit other people's opinions about products, news, and trends, and the concept of "influential person" emerges. An influential person (or *social media influencer)* today is considered a marketing strategy.

The Brazilian dairy industry has been standing out every year, and one of the promising areas is cheese production. The 2019 annual report<fn id="fn1">1 https://ablv.org.br/o-setor/relatorio-ablv/ </fn> by ABLV (Associação Brasileira da Indústria de Lácteos Longa Vida) indicates that there was an increase of 32% in liters of milk destined for cheese production in Brazil compared to 2009, which is greater than the percentage growth of milk UHT (26%).

Intending to collect information from social networks to find influential people, who appreciate artisanal cheeses, and who can influence potential new consumers, this work presents REDIC, a proposal for analysis, recommendation, and content propagation network, considering the Brazilian artisanal cheese market. REDIC classifies the user's content and interactions using ontologies and complex networks, deriving new relationships and allowing interconnecting information on different social networks.

REDIC was developed to support the market research of artisanal cheeses in a renowned Brazilian agribusiness institution. The results obtained through feasibility studies showed that the solution allows the search for communities of digital influencers who talk about artisanal cheeses and the dissemination of information on the network.

CCS CONCEPTS

• **Information systems**; • **Information retrieval**; • **Document representation; Ontologies**; • **Expert systems; Social networks Recommender systems**;

KEYWORDS

Recommendation System, Social Network Analysis, SNA, Ontology, Social Network

ACM Reference Format:
Nedson D. Soares, Regina Braga, José Maria N. David, Kennya B. Siqueira, Thallys S. Nogueira, Emerson W. Campos, Emerson A. P. Moraes, and Priscila V.Z.C. Goliatt. 2021. REDIC: Recommendation of Digital Influencers of Brazilian Artisanal Cheese: REDIC: Recomendação de Influenciadores Digitais do Queijo Artesanal Brasileiro. In *XVII Brazilian Symposium on Information Systems (SBSI 2021), June 07–10, 2021, Uberlândia, Brazil.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3466933.3466987

1 INTRODUCTION

Nowadays, advances in information technologies turn different online social networks (OSN), such as YouTube[1], Facebook[2], Instagram[3] and Twitter[4], increasingly popular among Internet users [1]. Due to this popularity, OSNs are becoming a critical channel for business and marketing [2].

From a structural point of view, OSNs are complex structures composed of nodes, usually representing users, and edges representing some form of interaction, collaboration, or influence between

SBSI 2021, June 07–10, 2021, Uberlândia, Brazil

ACM ISBN 978-1-4503-8491-9/21/06…$15.00
https://doi.org/10.1145/3466933.3466987

[1]https://www.youtube.com/
[2]https://www.facebook.com/
[3]https://www.instagram.com/
[4]https://www.twitter.com/

these users, allowing the sharing and dissemination of information [3].

The study of the different types of relationships between the nodes of OSNs is called Social Network Analysis (SNA) [4]. SNA is generally carried out to reveal information related to users and their interactions. An example of interaction between two Twitter users is the "mention", where a user mentions another user in his tweet. As a result, it provides human behavior knowledge related to specific contexts, such as consumption profile, political trends, among other contexts [4]. Among the main analysis types, we can highlight (i) detection of influential users, who are individuals with significant impact on their OSNs and who can influence others; and (ii) the detection of communities, which is the discovery of groups with similar characteristics in the network [5].

The analysis of the data generated by the OSNs can allow the discovery of implicit relationships, generating new information and allowing a new cycle of analysis. For these analyses to be carried out, it is necessary to specify models to correlate this data. The use of semantic models as ontologies [16] may be an option in this context since the generation of new implicit relationships and those declared generates new knowledge.

From a business point of view, OSNs present an excellent opportunity to reach a significant audience for market surveillance [6]. In this sense, the use of SNA contributes to information that can only be perceived from specific analyzes. For example, detecting the most influential user on the network can mean finding the user who can present the product or reverse the lack of information to help a company advertise its products [7]. Another example is the detection of communities that can be used to satisfy consumers of companies that use SNA focused on commercialized products [8].

One of the Brazilian sectors that can benefit from these analyzes is agribusiness. According to [11], agribusiness still lacks solutions that allow analyzes of the consumer market so that improvements can be applied (e.g., product adequacy and cost reduction). One sector that stands out is milk and dairy products. The Brazilian dairy industry is one of the most critical sectors of the country's food industry, losing in revenue only to the meat industry [12]. The dairy derivative that has shown a significant increase in consumption in Brazil is cheese [10]. The country has several regions producing artisanal cheese that maintain centuries-old traditions in the manufacture of these cheeses. According to the 2019 annual ABLV, there was an increase of 32% in the volume of milk used for cheese production in Brazil compared to 2009, which is greater than the percentage growth of the volume destined for UHT milk (26%). The Embrapa yearbook[5] points out that the production of artisanal cheeses reached millions in 2018. However, the sector lacks research on the consumer market for these products.

From the marketing perspective and the dairy sector's challenges, traditional market research methods take time and are sometimes incomplete and unrepresentative [12]. In this context, studies carried out at a renowned Brazilian agribusiness institution [11, 12] point to opportunities for new environments that can provide significant margins in the market. In this sense, we consider that OSNs can contribute to market surveillance in agribusiness. Using social data as the main elements of analysis, OSNs can promote the analysis of productive structures in the agribusiness sector [13].

This article aims to present a proposal for a consumer recommendation to support the artisanal cheese market, identifying influencers and user communities that address issues related to artisanal cheese in OSNs. To achieve this goal, we propose an automated solution for data analysis based on OSN content to extract, process, and recommend communities of influencers through ontologies and inference algorithms, complex network metrics, and community detection algorithms.

This work's main contribution is the specification of a recommendation proposal based on identifying influencer's communities in the dairy sector to assist in disseminating information about artisanal cheeses and presenting guidelines for the marketing practice for producers. In this article, the focus is on Twitter OSN because it is considered attractive from the point of view of the rapid dissemination of information. However, other OSNs and the combination of these can also be used.

The article is organized as follows: Section 2 presents the methodology used to conduct this research. Section 3 presents related works considering social networks and data recommendation analysis. Section 4 presents the proposed solution. Section 5 describes a feasibility study considering the artisanal cheese domain. Finally, Section 6 presents the final considerations.

2 DESIGN SCIENCE RESEARCH METHODOLOGY

To conduct this research, we used the Design Science Research methodology [22]. In the design science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the construction and application of the designed artifact. Artifact evaluation provides feedback information and a better understanding of the problem to improve product quality and the design process.

Based on design science research [22], the approach proposes the creation of a solution (**creation of an artifact**), called REDIC (REcommendation of Digital Influencers of Brazilian artisanal Cheese), to assist producers in identifying the most artisanal cheeses sought/cited, as well as their potential consumers (**problem domain**). REDIC includes analysis techniques, favoring the understanding of the data. It uses ontologies and complex network techniques to analyze data and recommend content about artisanal cheeses and consumers. The approach is innovative because it uses information discovered in processing inference algorithms, structural analysis of complex networks, and recommendation techniques to understand agribusiness data (artisanal cheese data) extracted from OSNs (**the creation of innovative and purposeful artifacts**). From the literature analysis, no similar works were found that present an approach that analyzes and visualizes the OSNs information related to agribusiness. This information aims to help identify consumers by considering and emphasizing the discovery of implicit information (**a problem space and a proposed or implemented mechanism to find an effective solution**).

To verify the feasibility of the proposal, two research questions were answered:

[5]https://www.embrapa.br/busca-de-noticias/-/noticia/36560390/anuario-do-leite-2018-e-lancado-na-agroleite

RQ1. **Which artisanal cheese is the most mentioned among communities generated from users who talk about artisanal cheese at OSN?** The purpose of this RQ is to answer specifically which artisanal cheese is most cited among users of an OSN. The identification of the most mentioned product among communities can help producers in decision making. Also, by identifying the most popular artisanal cheeses among users, we recommend publications on artisanal cheese according to the user's content.

RQ2. **Who are the biggest influencers of artisanal cheese at OSN?** The purpose of this RQ is to highlight in an OSN the most influential users in the dairy segment, more specifically in the production of artisanal cheeses. Thus, we can use these users as influencers who can promote artisanal cheeses. We can also provide evidence of how producers can use influential users' information as guidelines for developing new products, packaging, and prices, for example.

Unlike traditional marketing, the proposed solution is online and in real-time, which makes it easier to track whether a particular campaign is working for the product or not. We evaluated our proposal using Twitter data (**an assessment of the artifact is crucial**). The results point to the viability of the solution.

3 RELATED WORKS

Some studies discuss the main concepts of recommendation systems and SNA to detect and recommend influential communities and users in OSNs. User influence is defined in various ways on social networks, encompassing the degree of trust, user status, semantic web, and others. According to Trusov et al. [14], a users' number of activities influences his followers' behavior and decisions. Considering this premise, they proposed a method to determine the effect of the user's activity using his activity history.

Aimed to classify the textual content extracted from the OSNs, it is essential to obtain a precise understanding of the text's semantics [15]. In [17], a method based on ontologies was proposed, which included blog content such as authors, readers, and their interactions, to identify relevant topics. The identified topics were linked to the influential user, and decision-makers used the content posted for marketing policies. However, the study is limited to blog content. In [9], the authors developed the Short Semantic Pattern technique to collect semantic patterns in short texts. The technique extracted strings of words that share a similar meaning in Donald Trump's tweets.

In [18], the OLFinder algorithm is used to calculate the capacity and status of each Twitter user in a specific domain. From these calculations, a ranking list of influential users is obtained in the specific domain. In [19], the authors proposed grouping techniques in a researcher's social network to identify communities to generate recommendations for locations. In [20], the authors implemented the Coral platform, an SNA-based approach, to support business incubators' activities to improve the local community and their relationships.

Studies [9, 21] indicate that additional research is needed, as it is necessary to improve the understanding of purchasing behavior in agribusinesses, such as analyzing consumer groups and the consumer experience. However, currently, the primary method for collecting this information in agribusiness is questionnaire research [22], although some studies have already analyzed the use of the hashtag #farmersmarket to understand the opinions and experiences of social network users [6]. The results have practical implications for the management of agribusiness markets. However, users' demographic information is not analyzed due to the limitations of the application.

REDIC uses ontologies, SNA, and recommendation techniques to suggest classified content precisely focused on the artisanal cheese market compared to these works. The approach is innovative in the Brazilian artisanal cheese sector by using intelligent tools to analyze data on social networks to extract information and foster the market, for example, information on trends in the artisanal cheese market or consumer demographic information. The analyses performed generated useful results for marketing activities and consumer relationship management, and the methods employed may be applicable in other domains.

4 REDIC

To conduct the research, as previously stated, we use the design science research methodology [22]. First, we investigate the domain. Related works were analyzed to find similar proposals in the literature (Section 2). Based on this investigation, the REDIC solution was proposed. Therefore, the research contribution is the proposal of a solution that identifies and recommends influencers and published content related to artisanal cheese to collaborate with the sector's marketing strategies, using recommendation systems techniques. REDIC aims to implement subprocesses to detect published content and influential user communities on social networks to recommend new content and influencers of artisanal cheese.

4.1 Proposed Solution

The workflow in Figure 1 shows the sub-processes in which REDIC uses an OSN for data collection. This subprocess also extracts publications from its database. **Data collection and storage** begin with creating an artisanal cheese folksonomy[6] validated by dairy specialists (Figure 2-A and B). This activity aims to generate a set of keywords organized in categories that describe the artisanal cheese domain. The artisanal cheese folksonomy created is used to instantiate the ARTCHEE-O (ARTisanal CHEEse Ontology), detailed in Section 4.1.1.

The solution executes a data mining process (Figure 2-C), which extracts the ontology keywords to search for relevant publications in the OSNs. The collected publications are instantiated in the ontology, and the Pellet[7] inference algorithm is used to classify the data mined according to the content and category. A **cognitive analysis** algorithm, called Age Gender Estimation[8], estimates demographic data (e.g., age and gender) from the person's OSN profile image. With the data classified and instantiated in ARTCHEE-O, the **SNA** subprocess starts. In this sub-process, we use complex network theory to create a network of communities of digital influencers of artisanal cheese. SNA extracts from the ontology the

[6]Keywords that annotate and describe online content [24].
[7]https://github.com/pwin/owlready2/tree/master/pellet
[8]https://github.com/yu4u/age-gender-estimation

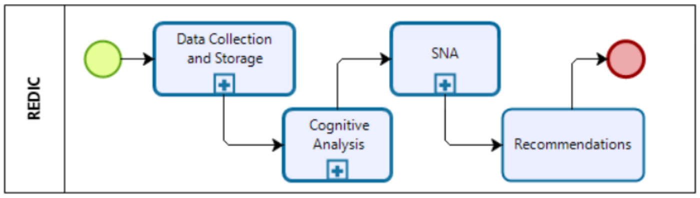

Figure 1: REDIC workflow.

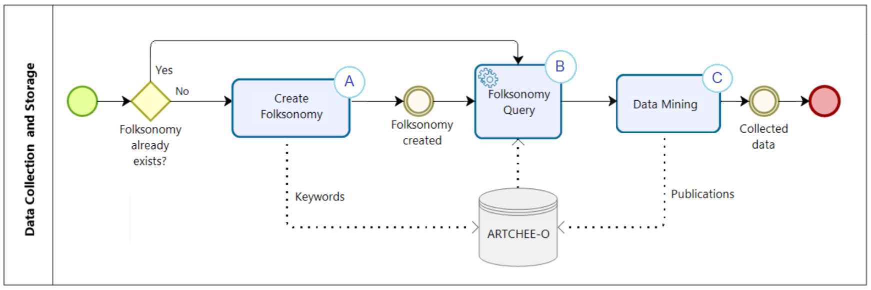

Figure 2: Data Collection and Storage Subprocess.

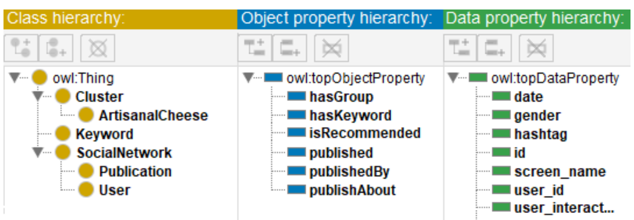

Figure 3: ARTCHEE-O's classes, *object properties,* and *data properties.*

instantiated publications. Then it extracts the interactions (e.g. mentions) that happen in the publications. We describe this step in more detail in Section 4.1.2; and finally, **recommendations** for content and grouped and classified influencers are made through database queries.

From the data generated by the previously described subprocesses, it is possible to query about influencers who talk about a type of artisanal cheese, specifying a gender, age and lifestyle.

4.1.1 ARTCHEE-O. The ARTCHEE-O ontology was specified, based on the folksonomy validated by specialists, to organize and infer knowledge about artisanal cheese. We developed the ontology to organize and extract relevant information using semantic rules.

Content and user resources are extracted from publications to be instantiated in the ontology, and, through the use of semantic rules (detailed next), new relationships are specified, generating implicit knowledge. Figure 3 presents the main elements of the ontology (*Cluster, Keyword, hasGroup, date*, among others), and we detail the main rules specified.

The ontology rules were developed using SWRL language[9]. We present some of the rules below:

[9]https://www.w3.org/Submission/SWRL/

Rule1: Publication(?tw) ^ swrlb:equal(?a, ?b) ^ hashtag(?k, ?a) ^ Keyword(?k) ^ hashtag(?tw, ?b) -> hasKeyword(?tw, ?k)

Rule2: Publication(?tw) ^ hasGroup(?k, ?gm) ^ hasKeyword(?tw, ?k) -> hasGroup(?tw, ?gm)

Rule3: Publication(?tw) ^ User(?us) ^ user_id(?us, ?a) ^ user_id(?tw, ?b) ^ swrlb:equal(?a, ?b) -> publishedBy(?tw, ?us)

Rule4: Publication(?tw) ^ hasGroup(?tw, ?gm) ^ publishedBy(?tw, ?us) -> publishAbout(?us, ?gm)

Rule5: User(?us) ^ publishAbout(?us, ?gm) ^ Publication(?tw) ^ hasGroup(?tw, ?gm) -> isRecommended(?us, ?tw)

First, the folksonomy is instantiated into the ontology by relating each keyword to its category through the "hasGroup" object property. After this stage, the process of storing publications in the ontology begins, referring to the instantiated folksonomy. **Rule1** relates publications to keywords. According to its data property "*hashtag*", the object property "*hasKeyword*" is inferred from the publication. **Rule2** relates the publication to the category. In **Rule3**, the publication of the object property "*publishedBy*" is inferred, indicating who owns it according to its data property "*user_id*". In **Rule4**, the object property "*publishAbout*" is inferred, which indicates the category of the user's post. Finally, **Rule5** recommends publications to the user, inferring the object property "*isRecommended*" according to the categories he published.

To instantiate ARTCHEE-O, we used the Python OwlReady2[10] library. The ontology can be accessed at[11].

4.1.2 SNA. From the data instantiated in ARTCHE-O, the activities responsible for modeling the user network begin (Figure 4-A and B). This subprocess will detect user communities that are interactively closer and quantify the influence of those users. For this, the NetworkX[12] library, a Python package, is used to create and

[10]https://github.com/pwin/owlready2
[11]https://github.com/nedsons/artchee-o
[12]https://github.com/networkx/networkx

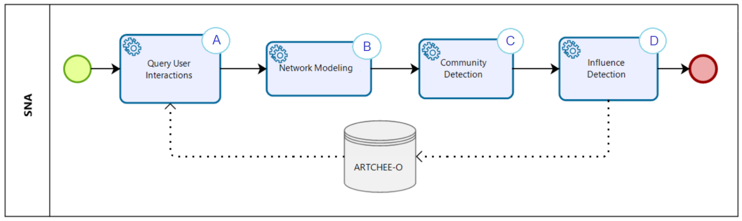

Figure 4: SNA Subprocess.

study the complex network structure, as is the case with a social network.

The result of the network modeling activity is a graph **G (V, E)** where: (i) a node $v_i \in \mathbf{V}$ indicates a user; (ii) an edge $e_{ij} \in \mathbf{E}$ indicates an interaction between the 'i' node to the 'j'. If any interaction between 'v_i' and 'v_j' users is found in the instantiated publications, an 'e_{ij}' edge without direction will be created to connect both nodes. In other words, if a user mentions another user.

In modeling, each edge has a weight that considers how much users appreciated the publication, such as the Twitter Likes[13] metric or Facebook Likes[14]. We are interested in describing the interaction between two users without focusing on their guidance for this analysis.

The community's detection is executed (Figure 4-C) that groups the network generated in different potentially overlapping communities, which are highly interconnected nodes. The Louvain algorithm is one of the most widely used algorithms for identifying communities due to its speed and high modularity [25]. The modularity values can vary from - 1 to 1, and the higher the value, the better the structure of the formed community[15].

Influence detection (Figure 4-D) aims to determine the user's influence using network metrics. A subgraph **SG** is extracted based on a set of nodes of the graph G. In our work, **SG** corresponds to the most populous community in the network. With generated **SG**, an analysis of the centrality of the network is performed. To visualize and quantify the user's relevance in the network, the Closeness (CC), Betweenness (BC), and the Degree of Centrality (DC) metrics are used.

We use the DC[16] measure, assuming that an important user has many connections. However, the calculation of this measure does not consider the overall structure of the network. For example, even if a node is connected to many others, it may not be a quick method of disseminating knowledge due to its position in the network. To consider the structure of the network, we propose the CC[17] measure. We use this measure to find users who are close to others. A limitation of this measure is that it is generally restricted to networks with related components since nodes that belong to different components do not have a path between them. However, this does not apply to our analysis, as we separate the network into connected communities.

Besides, we must consider that nodes that are not directly connected do not mean that they cannot interact on the network. They can communicate through other nodes. Thus, we use the BC[18] measure to find a user who acts as a bridge to other users. It's important to note that calculating the centrality betweenness of all nodes on a graph involves calculating the shortest paths between all pairs of nodes on a graph, which takes $O(|V||E| + |V|^2 \log |V|)$ time using Brandes' algorithm [27]. Thus, depending on the size of the network, it may not be feasible to use the BC metric. However, grouping the network into communities helps in this matter by reducing the number of nodes in the BC calculation.

5 FEASIBILITY STUDY

The Design Science methodology emphasizes the importance of an adequate evaluation. In [22], the authors state that the selection of evaluation methods must be appropriately combined with the projected artifact and the selected evaluation metrics.

Thus, to check the technical feasibility of the proposed solution, a feasibility study was carried out. The scope of the study was defined through the GQM (in English, Goal, Questions, Metrics): "**Analyze** the use of the REDIC solution **from the point of view of** researchers and consumers of artisanal cheeses **in the context of** data extracted from OSNs".

Twitter is an OSN considered attractive from the point of view of quick information dissemination. Also, it has much more public than protected profiles, and for those reasons, it is considered a target media for marketing [23]. In this sense, we use Twitter, as the target OSN in this study, processing the data collection through its application programming interface (API). The tweets are collected through the Tweepy[19] library, which was implemented using a list of keywords related to artisanal cheese as a filter.

Guided by research questions **RQ1** and **RQ2**, we collected a list of tweets using as a filter the list of keywords related to artisanal cheese, as detailed above. Using the ontology metadata, the result of the collection process was 6,543 tweets from Brazil's Portuguese language, tweeted in different parts of the world during the periods of 09/01/2020 to 11/06/2020. When we processed the Pellet algorithm on the collected data, the result obtained is shown: (i) in Figure 5, which presents an overview of inferred relationships focusing on the tweet_13 ontological individual. According to the ontological metadata, it was possible to classify the tweet in the

[13]https://help.twitter.com/en/using-twitter/liking-tweets-and-moments
[14]https://www.facebook.com/help/110920455663362?helpref=uf_permalink&rdrhc
[15]The activity used the Python Louvain Algorithm for Community Detection package to generate a resulting G-induced network, where the nodes are communities.
[16]It's an measure that individually quantifies how connected a node is in the network, using the number of direct connections of that node [21]
[17]Through this measure it is possible to quantify the shortest path from one node to another node through the network [21].
[18]This measure quantifies the degree to which a node is on the shortest path between two other nodes and is capable of acting as a bridge [21].
[19]https://www.tweepy.org/

Table 1: The 3 most cited kinds of cheese in the most populous community

Keyword	Category	Frequency
Queijo Coalho	Coalho	107
Queijo Canastra	Minas Artesanal	27
Queijo Meia Cura	Outros	22

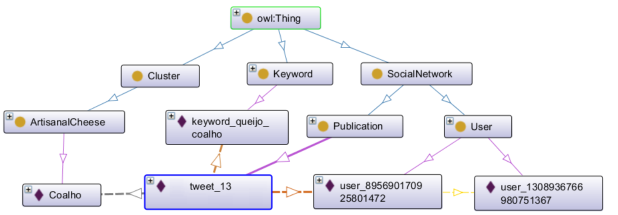

Figure 5: Relationship chart focusing on the individual "tweet_13".

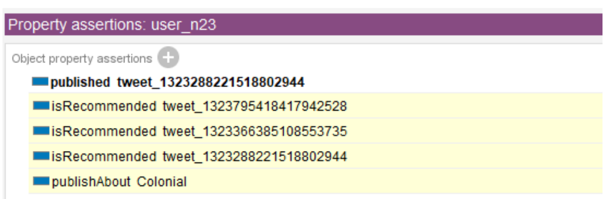

Figure 6: "user_23" *Object properties.*

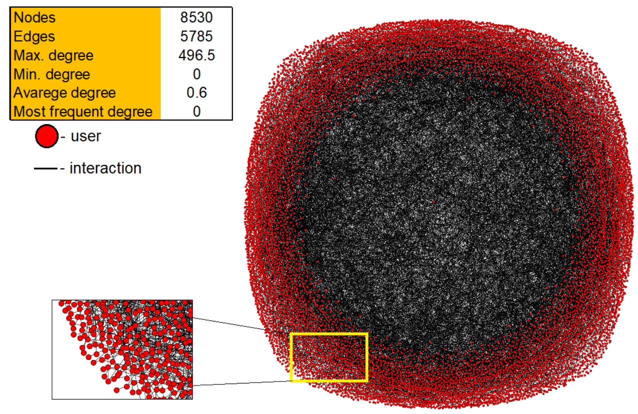

Figure 7: Generated G graph and general information.

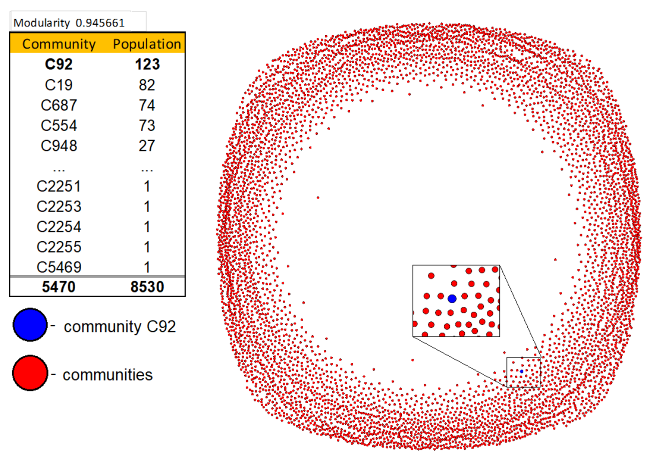

Figure 8: G induced community network and modularity.

category ***Coalho*** cheese; and (ii) in Figure 6, which shows the object properties assertions of *user_23* of the ontological *user* class. According to the metadata, the ontology recommended to this user, publications in the ***Colonial*** category.

Besides, the ontology also identified synonyms and spelling errors based on the specified rules. For example, if the keyword "**queijo coalho**" were misspelled for "**queijo coalio**", it would be possible to mix it by assigning several data properties hashtag to the individual keyword_queijo_coalho. Finally, we emphasize that although the ontological metadata of artisanal cheese must be predefined, publications and keywords can be instantiated automatically.

5.1 Results

Based on the results of the study, responses to research questions can be analyzed. Through the collected and analyzed content, it is possible to verify the technical feasibility of REDIC.

Considering the **RQ1** research question, interactions between users found in the collected tweets to generate the user network were extracted. An example of the generated **G** graph can be seen in Figure 7, along with its general information. With 8,530 nodes and 5,785 edges, the graph's minimum degree is 0, and it is a disconnected network containing nodes that have no edge.

Analyzing this number of nodes requires significant computational cost. Aimed to analyze only connected networks and reduce the computational cost of the analysis, the next process activity divided the graph into user communities as described in Section 4.1.2.

Figure 8 shows the induced graph where the nodes are communities. 5,470 communities were detected with modularity of 0.94.

Analyzing the most populous community, C92, with 123 users, we generated Table 1, which shows the three keywords most cited by users and their categories. It is observed that the most cited keyword is ***Queijo Coalho***, of the category ***Coalho***, with 107 citations, followed by ***Queijo Canastra*** and **Queijo Meia Cura**.

Based on the data generated and analyzed, the **RQ1**. "Which artisanal cheese is most mentioned among the communities generated from users who talk about the artisanal cheese at OSN?" could be answered: the visualization of users classified demographically and

B

Nodes	123
Edges	135
Max. degree	12.4
Min. degree	0.1
Avarege degree	0.6
Most frequent degree	0.2

C

Node	DC	BC	CC
V1	**1.0000**	**0.8802**	**0.4469**
V2	0.0732	0.2184	0.3653
V3	0.1138	0.1492	0.3536
V4	0.1138	0.1095	0.3262
V5	0.0569	0.1081	0.3211
...	...	...	...
V199	0.0081	0.0000	0.0000
V120	0.0244	0.0163	0.0163
V121	0.0569	0.0000	0.0000
V122	0.0244	0.0000	0.0000
V123	0.0081	0.0000	0.0000

- user

- interaction

*size of node correpond to CC metric

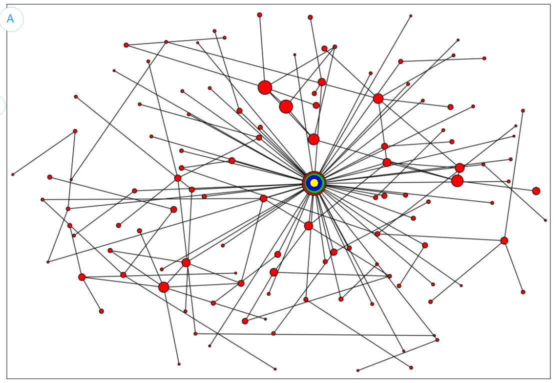

Figure 9: Community C92 and general information.

grouped according to their interactions can help researchers to identify consumers with similar characteristics. For example, instead of sending messages individually to each person related to their area of interest, gender, or age, a message directed to a community can be sent, or a product can be recommended to a community. Table 1 can be generated for each community, making it possible to search for the most cited artisanal cheeses per community. Thus, these data recommendation can assist the producer in meeting customer demands and monitoring the cheese market in OSNs.

Considering **RQ2**, we applied the influence analysis to the C92 community, which is the most populous. The SG92 subgraph corresponding to C92 can be seen in Figure 9-B, emphasizing the most influential node. The subgraph's general information can be seen in Figure 9-A, which, according to the minimum degree, the subgraph is a connected network. The result of the influence analysis is a list with the centrality score for each node. Figure 9-C shows a table with some of the users scored where the GC metric was normalized between 0 and 1 and, in bold, the node v1, which is the user that has the greatest potential of influence in the C92 community because the values of their metrics are higher when compared to other users.

Thus, **RQ2** ("Who are the leading influencers of artisanal cheese at OSN?") Could be answered: the identification of influencers in the users' network who talk about artisanal cheese on Twitter could be made through Figure 9, in which the user **v1** was the user of the most populous community with the highest score in the metrics used. Thus, decision-makers or producers can use this information to disseminate knowledge to public opinion or make a sale. When it comes to digital word of mouth marketing, this user can present artisanal cheese information or make product suggestions online. Figure 9 can be generated for the other communities, searching for each community's most significant influencer to choose the most convenient. In the end, the recommendation of the most influential user is made through queries to the ontology, visualization of graphs and tables. Besides, **v1** can use the recommended publications, such as the **user_n23** individual (Figure 6), to acquire more information. As a result, he does not need to visit several different OSN profiles to search for product information.

6 FINAL CONSIDERATIONS

This article presented a solution, called REDIC, to identify more relevant users and detect communities for marketing in the dairy sector, specifically artisanal cheese. The solution uses recommendation system techniques based on data mining, semantic web, and social network analysis. The solution was used on Twitter to verify the proposal's feasibility, considering comments from artisanal cheeses and their consumers. Thus, could use it to visualize the consumer cheese network structure separated into connected groups. Also, the identification of communities could help to discover relevant information, such as, for example, new communities of potential consumers.

From the proposed solution and the feasibility study, we could better understand the impact generated in the production routine of artisanal cheese producers, consumers, and researchers. As a result, it is possible to generate scientific knowledge about using the proposed solution to support agribusiness. REDIC as a marketing strategy and monitoring to solve the challenges of information systems in the field of agriculture, more specifically in the artisanal cheese producer business, can assist in the development of new products, insertion in markets, adding value, identifying weaknesses and opportunities, targeting marketing campaigns, among others

As future work, it is proposed to explore other domains and other techniques to support the recommendations. Geographic information and multimedia content can be explored to define intelligent user segmentation. It is also necessary to carry out a more comprehensive evaluation with other OSNs to verify the proposed approach's scalability. With the implementation of other OSNs, we intend to search for a connection between social networks using user's information as a parameter. Thus, we can disseminate information on an OSN through other OSNs. Additional metrics used in the SNA can also be studied.

REFERENCES

[1] Wesley Wing Hong Cheng, Ernest Tak Hei Lam and Dickson K.W. Chiu. 2020. Social media as a platform in academic library marketing: A comparative study. The Journal of Academic Librarianship, vol. 46, no. 5, (Sep, 2020). DOI: 10.1016/j.acalib.2020.102188.

[2] Ravneet Kaur and Sarbjeet Singh. 2017. A comparative analysis of structural graph metrics to identify anomalies in online social networks. Computers & Electrical Engineering, vol. 57, (Jan, 2017), 294–310. DOI:

10.1016/j.compeleceng.2016.11.018.
[3] Dariusz Król. 2014. On Modelling Social Propagation Phenomenon, in Intelligent Information and Database Systems. Springer International Publishing, 227–236.
[4] John Scott. 1988. Social network analysis. Sociology, vol. 22, no. 1, 109–127.
[5] Fabián Riquelme and Pablo González-Cantergiani. 2016. Measuring user influence on Twitter: A survey. Information processing & management, vol. 52, no. 5, 949–975.
[6] Ladislav Pilař, Tereza Balcarová, Stanislav Rojı́k, Ivana Tichá and Jana Poláková. 2018. Customer experience with farmers' markets: what hashtags can reveal. International Food and Agribusiness Management Review, vol. 21, no. 6, (Jul, 2018), 755–770. DOI: 10.22434/ifamr2017.0039.
[7] Youngsang Cho, Junseok Hwang and Daeho Lee. 2012. Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach. Technological Forecasting and Social Change, vol. 79, no. 1, (Jan, 2012), 97–106. DOI: 10.1016/j.techfore.2011.06.003.
[8] Francesco Bonchi, Carlos Castillo, Aristides Gionis and Alejandro Jaimes. 2011. Social Network Analysis and Mining for Business Applications, ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, (Apr, 2011), 1–37. DOI: 10.1145/1961189.1961194.
[9] Danielly Sorato and Renato Fileto. 2019. Linguistic Pattern Mining for Data Analysis in Microblog Texts using Word Embeddings. In Proceedings of the XV Brazilian Symposium on Information Systems - SBSI19. DOI: 10.1145/3330204.3330228.
[10] Kennya B. Siqueira. 2019. O mercado consumidor de leite e derivados, Embrapa Gado de Leite - Circular Técnica (infoteca-e), 1-17.
[11] Vinícius V. P. D. R. Nardy, Glauco R. Carvalho and Denis T. da Rocha. 2019. Mercado de leite fluido e queijos no Brasil: uma análise de 2005 a 2016. Embrapa Gado de Leite, 1-5
[12] Nedson D. Soares, Regina Braga, José Maria N. David, Kennya B. Siqueira, Victor Ströele and Fernanda Campos. Uma Arquitetura para a Recomendação de Consumidores de Queijo Artesanal Brasileiro. In Anais do Brazilian e-Science Workshop (BRESCI 2020). DOI: 10.5753/bresci.2020.11189.
[13] Edson Talamini and. Gabriel Murad V. Ferreira. 2010. Merging netchain and social network: Introducing the social netchain concept as an analytical framework in the agribusiness sector. African journal of business management, Academic Journals, 2981-2993.
[14] Michael Trusov, Anand V. Bodapati and Randolph E. Bucklin. 2010. Determining Influential Users in Internet Social Networks. Journal of Marketing Research, vol. 47, no. 4, (Aug, 2010), 643–658. DOI: 10.1509/jmkr.47.4.643.
[15] Bilal Abu-Salih, Pornpit Wongthongtham and Kit Yan Chan. 2018. Twitter mining for ontology-based domain discovery incorporating machine learning. Journal of Knowledge Management, vol. 22, no. 5, (Jun, 2018), 949–981. DOI: 10.1108/jkm-11-2016-0489.
[16] Nicola Guarino. 1998. Formal ontology in information systems. Proceedings of the first international conference (FOIS'98), Trento, Italy IOS press, 3-15.
[17] Feng Li and Timon C. Du. 2011. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. Decision Support Systems, vol. 51, no. 1, (Apr, 2011), 190–197. DOI: 10.1016/j.dss.2010.12.007.
[18] Abolfazl Aleahmad, Payam Karisani, Maseud Rahgozar and Farhad Oroumchian. 2016. OLFinder: Finding opinion leaders in online social networks. Journal of Information Science, vol. 42, no. 5, (Jul, 2016), 659–674. DOI: 10.1177/0165551515605217.
[19] Manh Cuong Pham, Matthias Jarke, Ralf Klamma and Yiwei Cao. 2011. A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis. Verlag der Technischen Universität Graz. DOI: 10.3217/JUCS-017-04-0583.
[20] Rafael Escalfoni, Mônica F. da Silva and Jonice Oliveira. 2020. Analyzing Social Relations in Startup Ecosystems. In XVI Brazilian Symposium on Information Systems. DOI: 10.1145/3411564.3411617.
[21] Loet Leydesdorff. 2007. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. Journal of the American Society for Information Science and Technology, vol. 58, no. 9, 1303–1319.
[22] Alan R. Hevner, Salvatore T. March, Jinsoo Park and Sudha Ram. 2004. Design science in information systems research. MIS quarterly, 75–105.
[23] Tiancheng Lou and Jie Tang. 2013. Mining structural hole spanners through information diffusion in social networks. In Proceedings of the 22nd international conference on World Wide Web - WWW 13. DOI: 10.1145/2488388.2488461.
[24] James Sinclair and Michael Cardew-Hall. 2007. The folksonomy tag cloud: when is it useful?, Journal of Information Science, vol. 34, no. 1, (May, 2007), 15–29. DOI: 10.1177/0165551506078083.
[25] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, (Oct, 2008). DOI: 10.1088/1742-5468/2008/10/p10008.
[26] Victor R. Basili, Gianluigi Caldiera and Dieter Rombach. 1994. The goal question metric approach. Encyclopedia of Software Engineering.
[27] Ulrik Brandes. 2001. A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2), 163-177.